A Systematic Review of Word Sense Disambiguation Systems: State of the Art Techniques and Challenges for Low-Resource Languages

Anonymous ACL submission

Abstract

Recent advances in Natural Language Processing (NLP) have been driven by the widespread adoption of Large Language Models (LLMs). Despite these improvements, state-of-the-art NLP models still struggle with ambiguous words, often failing to recognise the intended meaning of less commonly used terms in a sentence. This ambiguity problem impacts various linguistic tasks including machine translation and information retrieval, underscoring the importance of Word Sense Disambiguation (WSD). While significant progress has been made in WSD for high-resource languages like English, a notable research gap exists in understanding how current methods perform across multilingual and low-resource settings. Moreover, the impact and potential of LLMs in advancing WSD remain underexplored. This work presents a critical analysis of computational approaches to WSD, evaluating their effectiveness across English, multilingual, and low-resource contexts. We highlight current challenges for state-of-the-art systems and propose future directions in this evolving field.

1 Introduction

In Natural Language Processing (NLP) tasks, lexical ambiguity persists as a major challenge due to the multiple possible meanings of a word or phrase. Language understanding and language generation require effective algorithms to capture word senses in a sentence to perform multiple tasks like machine translation (MT), speech recognition (SR), and information retrieval (IR) (Pu et al., 2018; Zong et al., 2022; Goldwater et al., 2010). Misinterpreting a word's sense in tasks like MT and IR, particularly when embedded in social media interfaces, can lead to misinformation (Carpuat and Wu, 2007). Ambiguous words, which are commonly used in day-to-day communication, are particularly challenging to interpret automatically in low-resource language situations.

Often part of an NLP pipeline, Word Sense Disambiguation (WSD) aims to disambiguate the correct sense of a word from a sense space by analysing the linguistic features of the sentence and the words around the ambiguous word. Ambiguity in natural text can be categorised into four groups: lexical, syntactic, semantic, and pragmatic. Syntactic ambiguity concerns sentence structure based on phrase structure, coordination, modification, and scope (MacDonald et al., 1994), while pragmatic ambiguity involves unclear intent based on context (Macagno and Bigi, 2018). Semantic ambiguity refers to multiple meanings at the language levels of phrases, sentences, or entire passages, while lexical ambiguity specifically concerns individual words with multiple meanings. This work focuses on various aspects of lexical ambiguity and computational approaches to addressing this issue.

While several reviews on WSD exist, there is no systematic review of it in multilingual and low-resource language settings (Nanjundan and Mathews, 2023; Bevilacqua et al., 2021), nor of the effect of the integration of modern approaches like Large Language Models (LLMs). Through a systematic literature analysis, this study identifies critical limitations in current WSD research and outlines potential future directions. This work contributes a comparative examination of WSD algorithms across English, multilingual, and low-resource contexts. It uniquely incorporates LLMs into the discussion. Next, we present a comprehensive review of WSD research, covering both classical approaches and modern LLM-based methods, followed by an analysis of persistent challenges in WSD and proposed future directions.

2 Background

Previous research shows that words require the use of their context to resolve their meaning, showing that isolated word analysis can be prone to errors (Luo et al., 2018a). For instance, "*I connected*

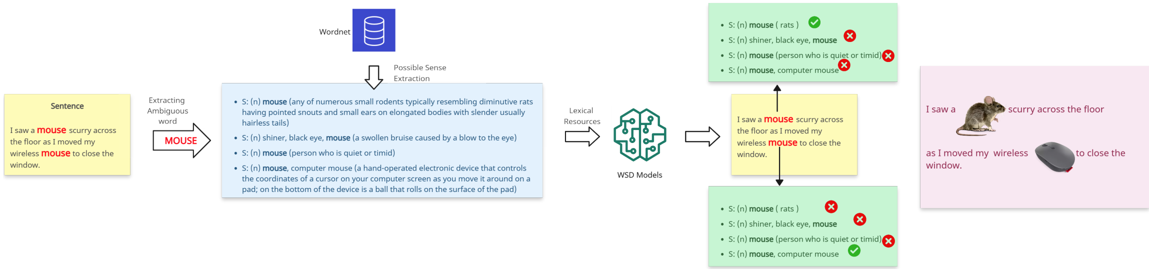

Figure 1: Disambiguation process for the word "Mouse"

the wireless mouse for easy navigation." contains *mouse* as an ambiguous word. According to WordNet[1], the word "mouse" contains four different noun meanings, including "*a small rodent typically resembling diminutive rats*" and "*a hand-operated electronic device that controls the coordinates of a cursor*". However, identifying the correct sense of this problem depends on the context around the particular usage of "mouse". The word "wireless" suggests an electronic device, which indicates that "mouse" here more likely refers to an electronic device rather than a rat or silent person. Figure 1 demonstrates the disambiguation process of the word "mouse" in an ambiguous sentence. Therefore, positional features, Part-Of-Speech (POS) tags, elements of the entire sentence, and context around the word for disambiguation play a prominent role in WSD. These parameters can be efficient in identifying lexical ambiguity in natural text, while glosses from WordNet and lexical knowledge like synonyms, hypernyms, and antonyms can be used to improve the disambiguation process.

2.1 Evolution of WSD Techniques

WSD research began in the 1950s with rule-based methods using hand-crafted rules based on context-based clues like grammatical and syntactic structures (Bowerman, 1978). These were labour-intensive and lacked scalability, as complex languages with more vocabulary require a vast number of rules (Palmer et al., 2006). In the 1980s, knowledge-based approaches emerged using lexical resources like WordNet, dictionaries and Thesauri and algorithms like Lesk (Miller, 1992; Lesk, 1986) to calculate overlapping words between the context and the dictionary definitions, though they struggled with unseen senses. In the 1990s, researchers explored supervised machine learning methods using annotated corpora and algorithms like Naive Bayes, Decision Trees and SVM classifiers (Le and Shimazu, 2004; Al-Bayaty and Joshi, 2016; Gosal, 2015), limited by data scarcity and poor domain generalisation. To address this weakness, in the mid-2000s, semi/unsupervised techniques based on clustering and algorithms like Expectation Maximisation (EM) and Latent Semantic Analysis (LSA) (Pedersen, 2006; Dempster et al., 1977; Deerwester et al., 1990), using vector representations of words based on co-occurrence statistics were used. However, these approaches were relatively lower in accuracy due to noisy clustering. With the advent of modern neural network architectures, Word embeddings (e.g., word2vec, GloVe, ELMo) improved word representation, producing smaller dense vectors to capture the semantic similarity based on target and context word co-occurrence patterns (Church, 2017; Pennington et al., 2014; Kutuzov and Kuzmenko, 2019). However, static vectors lacked dynamic sense disambiguation. Building on recent advances in machine learning algorithms, Transformer-based models like BERT and GPT (Vaswani, 2017; Huang et al., 2020) have advanced WSD with contextual embeddings and domain generalisation. Current trends explore Few-shot/Zero-shot learning and in-context learning using LLMs (Basile et al., 2025; Yae et al., 2025).

2.2 Key challenges in WSD

Current WSD algorithms face significant challenges when applied to real-world data (Tyagi et al., 2022). High polysemy (multiple meanings) of ambiguous words poses a substantial challenge, particularly with less frequently used words. This remains a common issue for most systems, as training data contains limited examples of less frequent senses (Sumanathilaka et al., 2024c). Data scarcity, especially the lack of annotated corpora, remains a significant obstacle for modern WSD. High-quality, sense-tagged corpora are crucial for supervised

[1]A large lexical database of English developed at Princeton University (Miller, 1995)

learning approaches, but creating large annotated datasets is time-consuming and expensive. This problem is particularly serious for low-resource languages, making training high-performing models exceptionally challenging. Although automated approaches have been used to generate synthetic data, these methods have failed to produce sufficiently diverse corpora to handle less frequent senses (Ganesh et al., 2024). Furthermore, due to processing window limitations, classical models struggle with the contextual complexity of multi-sentence contexts (Loureiro et al., 2020). Yet disambiguating senses often depends on long-distance and inter-sentential relationships. Certain domains, such as biomedical (Mondal et al., 2015), legal (Gliozzo et al., 2004), financial (Hogenboom et al., 2021), geospatial, and environmental, face significant ambiguity challenges due to the overlap of technical terms with general language (Buitelaar et al., 2006). To address these issues, researchers have increasingly focused on developing domain-specific models and enhanced disambiguation techniques that better account for specialised vocabulary nuances and contextual dependencies.

3 Methodology

This systematic review followed a PRISMA process (Page et al., 2021) for paper collection and analysis. The detailed methodology is provided in Figure 2. Our primary data sources included IEEE Xplore, the ACL Anthology, arXiv, SpringerLink, and the ACM Digital Library. Additionally, we utilised Google Scholar, Semantic Scholar, and ResearchGate to conduct supplementary searches and assess publication relevance for inclusion in this study. We employed 42 keywords related to WSD and 10 keywords related to general ambiguity (see Appendix A) when searching data sources within the timeframe of 1995 to mid-2025.

3.1 Pipeline for Data collection

This section presents the data collection pipeline, which incorporates human review and LLM-assisted analysis, specifically using GPT and Notebooklm to extract and filter relevant information from selected papers during the initial screening phase. We began by gathering papers from data sources using a basic set of keywords to establish filtering criteria. The first author then reviewed and annotated the selected articles with appropriate labels provided in Appendix A. Quality assessment was performed based on the Critical Appraisal Skills Programme checklist, examining the clarity of objectives, appropriateness of methodology, rigour of analysis, and relevance to WSD (Treloar et al., 2000). The following criteria were used to filter papers for the final study.

- Does the paper propose a novel technique?
- Does the paper explain the training and testing data used for the study?
- Does the paper report the final results of the study?
- For review papers, does the paper critically evaluate the approaches?
- For dataset creation papers, is the data creation process and data distribution discussed?

Out of the initially selected papers (213), 15 papers on English WSD were chosen for the meta-analysis based on their significant contributions to English WSD systems and the reported performance outcomes. The remaining papers (196) were reviewed and discussed within the article but were not included in the meta-analysis.

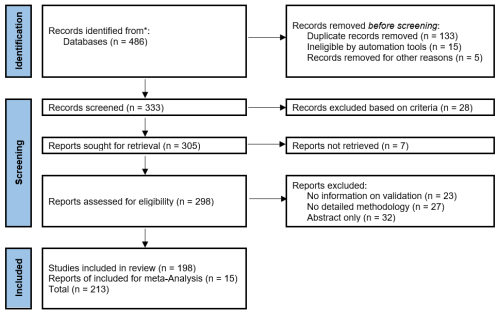

Figure 2: PRISMA process used for paper selection.

4 Approaches to WSD

In this section, we discuss the current approaches used in WSD, focusing on both English and multilingual approaches, including those for low-resource languages. Given their new status as a mainstream component of NLP, we will explore how LLMs are applied for WSD. Additionally, Tables 1 and 3 present a meta-analysis of different WSD methodologies across languages, providing a comparative overview of their performance and application. The datasets used in these studies are summarised in Table 4.

4.1 Knowledge-based Approaches

Knowledge-based (KB) WSD methods leverage external resources like lexical databases and ontologies to disambiguate word meanings. These methods utilise semantic similarity metrics and graph-based algorithms. Techniques like the Lesk algorithm and LSA have been employed for WSD. For example, Wang et al. (2020) used semantic space and paths within sentences to enhance WSD using WordNet. Chaplot and Salakhutdinov (2018) scaled words in context linearly using topic modelling, proposing a variant of Latent Dirichlet Allocation (LDA) with synset proportions. Various modifications of the initial Lesk algorithm (Lesk, 1986) have been proposed, such as (Agirre and Soroa, 2009), adapted Lesk (Banerjee et al., 2003), and enhanced Lesk (Basile et al., 2014). Lesk has been widely applied to low-resource WSD, particularly in Marathi (Gahankari et al., 2023a), Assamese (Gogoi and Baruah, 2022), Manipuri (Singh and Devi, 2024), Nepali (Singh et al., 2021) and Sinhala (Arukgoda et al., 2014). Graph-based algorithms are also prevalent in WSD. The Babelfy study, which connects Entity Linking (EL) to named entities, introduces a unified graph-based method for EL and WSD. This approach identifies potential meanings and selects the most coherent semantic interpretations using the densest subgraph heuristic, demonstrating effectiveness in multilingual settings (Moro et al., 2014). Early WSD solutions involved random walks over large KB like extended WordNet (Agirre et al., 2014). Jha et al. (2023c) utilised Hindi WordNet with weighted graphs representing word senses and their relations, while Duarte et al. (2021) combined graph-based approaches with word embeddings and contextual information for semi-supervised WSD. Exploiting WordNet relations, mainly Synset definitions, the Hypernymy relation, and definitions of context features, further enhances WSD accuracy (Kolte and Bhirud, 2009). Bootstrapping techniques incorporating WordNet synsets were employed by Gahankari et al. (2023b), and adaptive complex networks based on semantic similarities were explored for ambiguity resolution (Kokane et al., 2021). Martinez-Gil (2023a) emphasised the significance of contextual information by incorporating similarity measurements. The Synset Relation-Enhanced Framework (SREF) for enriched sense embeddings expanded the WSD toolkit by augmenting basic sense embeddings with sense relations and a try-again mechanism (Wang and Wang, 2020). Combining knowledge resources, such as cross-lingual approaches and KB models, has also shown promise. KB approaches have been applied in both English and other languages, such as the Persian WSD technique by Rouhizadeh et al. (2020) and cross-lingual approaches highlighted by Rudnick (2018). ExtEnD, which frames the task as a text extraction problem, leveraged transformer-based architectures to improve disambiguation accuracy (Barba et al., 2022). Semi-supervised WSD using graph-based SSL algorithms and various word embeddings combined with POS tags and word context were explored (Duarte et al., 2021). Studies such as Sumanathilaka et al. (2024c, 2023) proposed a suggestion-level module with a tree structure for Romanised Sinhala word disambiguation. In contrast, Perera et al. (2025) proposed a hybrid approach for Sinhala disambiguation, highlighting the value of KB models. These studies demonstrate the potential of KB, such as WordNet and BabelNet, which use semantic relationships to improve disambiguation. Recent work combine KB methods with machine learning to enhance scalability and adaptability. However, challenges remain in maintaining coverage for low-resource languages and handling ambiguity in dynamic, real-world contexts.

4.2 Supervised Learning Approaches

Supervised WSD is a well-researched field that relies on labelled datasets like Semcor and FEWS, along with WordNet, for training models (Scarlini et al., 2020a). Researchers have explored various computational techniques to address the WSD problem, often reframing it as a different computational challenge. For instance, ConSec introduced a novel approach by treating WSD as a text extraction problem (Barba et al., 2021b), incorporating a feedback loop to focus on the ambiguous word and its context. Song et al. (2021) enhanced sense interpretation by leveraging synonyms and example phrases, demonstrating the value of word senses and their glosses. The ESC approach reframed WSD as a span extraction problem using the ESCHER transformer-based architecture (Barba et al., 2021a), mitigating training data bias and achieving promising results. EWISER explored the integration of Lexical Knowledge Bases (LKB) by utilising synset embeddings and relations to train neural architectures (Bevilacqua et al., 2020), yielding positive outcomes in English WSD. Additional techniques like SpareLLM's use of sparse contextualised word representations (Berend, 2020) and the Bi-Encoder model's integration of target words with context and glosses (Blevins and Zettlemoyer, 2020) have been investigated. Various BERT variations, including fine-tuning pre-trained models, have also been explored in the context of WSD (Huang et al., 2020). Luo et al. (2018a) introduced a gloss-augmented WSD neural network that jointly encodes the context and glosses of the target word to model the semantic relationship between them within an improved memory network framework. This work has extended its gloss via semantic relations to enrich the gloss information using WordNet. Other approaches include context-dependent methods (Koppula et al., 2021), multiple sense identification (Orlando et al., 2021), and stacked bidirectional LSTM with attention mechanisms (Laatar et al., 2023). Data augmentation techniques like SMSmix have increased the frequency of Least Frequency Senses (LFS) (Yoon et al., 2022), addressing training data distribution bias. Kaddoura and Nassar (2024) uses an Enhancedbert for Arabic disambiguation, which aims to disambiguate 100 polysemous words, while El-Razzaz et al. (2021) used a BERT for Arabic WSD. Previous studies have highlighted the significant impact of synonyms and contextual meaning (paradigmatic relations) on sense identification.

4.3 Unsupervised Learning Approaches

Unsupervised WSD aims to determine the correct gloss of a word in context without using labelled (manually annotated) data. These methods explore patterns, distributions, and relationships in large, unlabelled text corpora to infer the correct sense (Mao et al., 2024). LSA, Latent Dirichlet Allocation (LDA), Page Rank approaches and clustering of words into distinct usages are prominently used (Vidhu Bhala and Abirami, 2014; Masethe et al., 2024). Knowledge-based unsupervised algorithms use lexical contents from knowledge resources to improve the quality of information required for the disambiguation process (Hogan et al., 2021). Niu et al. (2004) explored a context clustering scheme with a Bayesian framework to capture sense distinctions at the category level, allowing for WSD across the entire vocabulary with minimal annotated training data, ultimately outperforming existing unsupervised WSD systems. The incremental cluster-based graph structures proposed by Widdows and Dorow (2002) focused on the symmetric relationship between pairs of nouns which occur together. Jain and Lobiyal (2020) introduced an algorithm to identify hidden information connecting words in a sentence. This implicit information is represented through a graph, which aids in resolving word ambiguities in homonyms and polysemous words. Lin and Verspoor (2008) proposed a framework that incorporates semantic information into language models, enabling systems to address NLP tasks by combining syntactic and semantic cues. Chen et al. (2009) introduced a graph-based method for large-scale WSD, framing the task as identifying the most significant node (representing word senses) within a graph. Extending this approach, Jha et al. (2023b) employed weighted graphs where edges represent semantic relationships between word senses, enhanced by Hindi WordNet-based similarity weights, to improve sense selection using a random-walk algorithm. Başkaya and Jurgens (2016) developed a semi-supervised WSD system that combines limited sense-annotated data with sense induction techniques to automatically discern word meanings, outperforming purely supervised models on similar data. Sankar et al. (2016) introduced an unsupervised WSD model leveraging seed sets and collocations from a Malayalam corpus; the method expands initial seed sets to create sense clusters that identify the target word's meaning based on context. ShotgunWSD (Butnaru et al., 2017), performed document-level WSD in three phases: brute-force WSD for probable sense configurations, prefix and suffix matching, and ranking by length majority voting scheme based on the top configurations. This work was extended in Butnaru and Ionescu (2019) by introducing a relatedness score between word senses. Context-aware WSD systems, such as Martinez-Gil (2023b), allow for flexible integration of contextual cues in similarity measures. Wiedemann et al. (2019) focused on nearest-neighbour classification using contextual word embeddings (CWEs) and cosine distance, demonstrating the efficacy of distance-based approaches in WSD tasks. Recent studies explored cluster discrimination analysis over the semantic network with group algebra, noting considerable accuracy while reducing the parameter count (Guzman-Olivares et al., 2025). Padwad et al. (2024) proposed a BERT-based model supported by Euclidean distance between synsets for Hindi disambiguation, while Jha et al. (2023a) introduced a graph-based WSD algorithm for Hindi. Hou et al. (2020); Lyu and Mo

(2023) for Chinese WSD, and Djaidri et al. (2023); Alian and Awajan (2020) for Arabic WSD represent other notable WSD algorithms. However, these systems often face significant performance challenges, primarily due to the scarcity of lexical resources and the limitations of embedding-based synset representations, which can be noisy and hinder precise sense discrimination. The situation is further complicated by the morphological richness of languages such as Hindi, Sinhala, and Arabic, which adds complexity to clustering and graph-based methods. Despite these challenges, unsupervised approaches have shown particular promise for addressing WSD in languages with limited annotated data. To overcome data scarcity, many studies have constructed custom corpora or adapted existing WordNet resources for low-resource languages. Additionally, some cross-lingual strategies have leveraged resources from high-resource languages like English or utilised domain-specific corpora from sources such as Wikipedia, Twitter, and clinical notes (Jaber and Martínez, 2021).

4.4 Advances with Large Language Models

Recent advances in LLMs have led to a surge of interest in exploring their capabilities across various NLP tasks. Their unsupervised training on massive datasets has significantly enhanced their performance in language comprehension tasks. Sainz et al. (2023) demonstrated that LLMs possess an inherent understanding of word senses, suggesting their potential for WSD without specific training. They framed WSD as a textual entailment problem, prompting LLMs to assess the suitability of a domain label for a sentence containing an ambiguous word. Notably, this zero-shot approach outperformed random guessing and, in some cases, matched or even surpassed the performance of supervised WSD systems (Ortega-Martín et al., 2023). Additionally, cross-lingual word sense evaluation using contextual word-level translation on pre-trained language models has been explored, and zero-shot WSD has been evaluated through cross-lingual knowledge (Kang et al., 2023a). A contrastive self-training framework, COSINE, which fine-tunes pre-trained LLMs with weak supervision and no labelled data, was further investigated (Yu et al., 2021). Manjavacas and Fonteyn (2022) explored the use of non-parametric learning with effective fine-tuning of LLMs for Dutch and English historical resources. Qorib et al. (2024) demonstrated the effectiveness of encoder-only models compared to decoder-only models, while Sumanathilaka et al. (2024a) showed the effectiveness of prompt engineering techniques for WSD using in-context learning with GPT 3.5 Turbo and GPT 4. In their further studies, they benchmarked different LLMs' behaviour for WSD, showing that Deepseek-R1 and o4-mini are more suitable for disambiguation tasks compared to other flagship LLMs (Sumanathilaka et al., 2024d). These findings have been further supported by Kibria et al. (2024). Yae et al. (2024) discusses the relationship between LLM model sizes and WSD performance, while Cahyawijaya et al. (2024) showed the limitations in cross-lingual WSD in LLMs with false friends words[2]. Recent studies have discovered the pros and cons of large language models in both English and multilingual WSD, opening a new arena in efficient WSD (Kang et al., 2023b; David et al., 2024; Ren et al., 2024; Abdel-Salam, 2024; Laba et al., 2023). However, research on WSD for low-resourced languages remains under-explored due to the limited language support offered by current LLMs.

4.5 Discussion on Meta Review

Tables 1 and 3 present a chronological progression of WSD research from 2018 to 2024 and offer a comparative analysis of techniques. Early works such as GAS and EWISE employed hybrid, BiLSTM-based models that utilised gloss embeddings enriched with semantic relations. However, these models were inherently limited in their ability to model long-range dependencies. Moreover, their integration of external knowledge was not fully contextualised or dynamically applied, which constrained their overall effectiveness. From 2019 onwards, there has been a clear shift towards transformer-based architectures, primarily leveraging BERT and its variants. BERT's bidirectional self-attention mechanism allows for the deep contextualisation of word meaning based on both left and right contexts, which is crucial for distinguishing fine-grained word senses. Its ability to process entire sequences simultaneously rather than step-by-step, as in RNNs, enables more effective representation of polysemous words in context, as proven in GlossBERT and SenseEMBERT. Additionally, BERT and its variations use subword to-

[2]Orthographically similar but have completely different meanings

Models	Dev	Unified Eval Framework				POS Tag based UEF					FEWS Fewshot	
	SE07	SE2	SE3	SE13	SE15	N	V	A	R	ALL	Dev	Test
MFS	54.5	65.6	66.0	63.8	67.1	67.7	49.8	73.1	80.5	65.5	52.8	51.5
Lesk	51.6	63.0	63.7	66.2	64.6	69.8	51.2	51.7	80.6	63.7	42.5	40.9
Babelfy (Moro et al., 2014)	51.6	67.0	63.5	66.4	70.3	68.9	50.7	73.2	79.8	66.4	-	-
GAS (Luo et al., 2018b)	-	72.2	70.5	67.2	72.6	72.2	57.7	76.6	85.0	70.6	-	-
EWISE (Kumar et al., 2019)	67.3	73.8	71.1	69.4	74.5	74.0	60.2	78.0	82.1	71.8	-	-
(Vial et al., 2019)	69.5	77.5	77.4	76.0	78.3	79.6	65.9	79.5	85.5	76.0	-	-
LMMS (Loureiro and Jorge, 2019)	68.1	76.3	75.6	75.1	77.0	78.0	64.0	80.7	83.5	75.4	-	-
GlossBERT (Huang et al., 2020)	72.5	77.7	75.2	76.1	80.4	79.8	67.1	79.6	87.4	77.0	-	-
ARES (Scarlini et al., 2020d)	71.0	78.0	77.1	77.3	83.2	80.6	68.3	80.5	83.5	77.9	-	-
SenseEmBERT (Scarlini et al., 2020c)	60.2	72.2	69.9	78.7	75.0	80.5	50.3	74.3	80.9	72.8	-	-
EWISER (Bevilacqua et al., 2020)	71.0	78.9	78.4	78.9	79.3	81.7	66.3	81.2	85.8	78.3	-	-
SemEq $_{\text{Base Expert}}$ (Yao et al., 2021)	74.1	81.0	78.5	79.9	82.6	82.5	69.9	82.5	88.4	79.9	80.4	80.1
SemEq $_{\text{Large Expert}}$ (Yao et al., 2021)	74.9	81.8	79.6	81.2	81.8	83.2	71.1	83.2	87.9	80.7	81.8	82.3
ESR $_{\text{base}}$ (Song et al., 2021)	77.4	81.4	78.0	81.5	83.9	83.1	71.1	83.6	87.5	80.7	77.9	77.8
ESR $_{\text{Large}}$ (Song et al., 2021)	78.5	82.5	80.2	82.3	85.3	84.4	73.0	74.4	88.0	82.0	83.8	83.4
CoNSEC (Barba et al., 2021b)	77.4	82.3	79.9	83.2	85.2	85.4	70.8	84.0	87.3	82.0	-	-
SACE$_{\text{large}}$ (Wang and Wang, 2021)	82.4	81.1	76.3	82.5	83.7	81.9	84.1	72.2	86.4	89.0	-	-
ESCHER (Barba et al., 2021a)	76.3	81.7	77.8	82.2	83.2	83.9	69.3	83.8	86.7	80.7	-	-
RTWE $_{\text{Base}}$ (Zhang et al., 2023)	74.5	82.3	80.9	81.8	83.7	83.3	72.2	87.4	87.6	81.6	-	-
RTWE $_{\text{large}}$ (Zhang et al., 2023)	77.1	85.2	83.3	83.8	86.3	85.7	75.1	90.6	88.7	84.6	-	78.4
GlossGPT (Sumanathilaka et al., 2025)	76.2	86.1	82.9	75.4	83.0	82.6	73.1	91.9	88.6	81.8	90.2	90.7

Table 1: F1 score presented for flagship models using Semcor training data and FEWS

kenisation (WordPiece) to help handle rare and morphologically complex words, which often pose challenges in WSD. Techniques such as sentence pair classification in ESR, gloss alignment in SEMEQ, and similarity-based approaches combined with retry mechanisms in SACE have demonstrated strong performance, validating the capabilities of transformer-based models in capturing nuanced word senses. Recent advancements like GlossGPT highlight the field's evolution toward few-shot learning and chain-of-thought (CoT) prompting, aligning with the broader trend in LLMs. Throughout this meta-analysis, it is notable that verbs significantly underperformed compared to nouns and adjective disambiguation. This demonstrates the importance of focusing on verb disambiguation as a priority area for future WSD research, particularly through context-sensitive architectures and targeted lexical resources.

5 Evaluation Metrics

The effectiveness of a WSD algorithm is crucial, and evaluating such algorithms requires consistent metrics and computational techniques to benchmark performance (Zhang et al., 2025). The SemEval datasets (formerly known as Senseval) are indispensable for benchmarking and consist of multiple WSD-related tasks spanning different years. Senseval-2 (Edmonds and Cotton, 2001), Senseval-3 (Litkowski, 2004; Snyder and Palmer, 2004), SemEval-2007 (Pradhan et al., 2007a), 2013 (Navigli et al., 2013), and 2015 (Moro and Navigli, 2015) have established themselves as a standard in testing and comparing WSD systems across different time periods and paradigms (Raganato et al., 2017). FEWS evaluation set (Blevins et al., 2021) contains zero-shot and FEW-shot dev and test data, each with 5000 datatuples. The F1 score remains the most commonly used metric in WSD evaluation due to its balanced assessment of precision and recall. Also, accuracy, precision, and recall are frequently employed to provide a holistic view of system performance. In setups with highly diverse senses, evaluations at both the sense and word levels are often conducted to capture the nuanced behaviours in handling ambiguity. Table 1 presents the F1 score of the meta-analysis. While metrics such as F1 score and accuracy are integral to benchmarking, human evaluation provides invaluable insights into the efficacy of WSD systems. Human evaluators can assess both correctness and the appropriateness of senses in nuanced, context-dependent scenarios where automated systems might falter (Plaza et al., 2011). This process often involves linguists or domain experts who annotate datasets with sense labels, serving as the gold standard for evaluating system outputs. Human evaluation also facilitates error analysis, highlighting cases where systems misinterpret polysemy, metonymy, or rare senses (Murray and Green, 2004; Aimelet et al., 1999). Incorporating human evaluation alongside automated metrics offers a more comprehensive understanding of system performance, fostering advancements in algorithmic approaches and resource development for WSD.

6 Applications of WSD

Word sense disambiguation has been a major necessity in many computational linguistics tasks. WSD has been explored in many domains, and ablation studies have been conducted in the last few decades. WSD has been a key component in machine translation, and several key experiments have been considered. Chan et al. (2007) explored phrase-based MT integrating WSD systems, while Carpuat and Wu (2007) integrated WSD with statistical machine translation systems to enhance MT accuracy. Neale et al. (2016) demonstrated that word sense awareness is essential for accurate translation, while Wang et al. (2023) discussed the importance of WSD in Neural MT. Additionally, studies such as Carpuat and Wu (2005); Xiong and Zhang (2014); Costa-Jussá and Farrús (2014) explored statistical MT with WSD, and Pu et al. (2018); Iyer et al. (2023); Han et al. (2019) focused on Neural MT with WSD. Not only for MT, but IR has also significantly benefited from effective WSD. Zhong and Ng (2012) explored the use of word sense in language modelling approaches for IR, and Hristea and Colhon (2020) showed that the usage of WSD for unsupervised solutions in IR is impactful. Moreover, the SemEval-2007 Task (Agirre et al., 2008) investigated WSD in cross-lingual IR, further explored by Kang et al. (2004), Clough and Stevenson (2004), and Manwar et al. (2024). However, in linguistics-related tasks and applications, the involvement of WSD has been extensively discussed. For co-reference resolution, WSD is instrumental in resolving references to entities in texts with ambiguous terms (Sukthanker et al., 2020). It also supports morphological analysis by determining the correct sense of inflected or derived forms in morphologically rich languages. In Named Entity Recognition (NER), WSD helps distinguish polysemous terms, enabling the differentiation between named entities and common nouns or verbs (Garla and Brandt, 2013; Aliwy et al., 2021). Additionally, in lexical substitution, WSD ensures that word substitutions retain their original sense, enhancing paraphrase generation. Moreover, lexical chaining, semantic similarity computation, and knowledge-based reasoning rely on accurate WSD to avoid semantic inconsistencies and errors (Urena-López et al., 2001). These applications underscore the fundamental importance of WSD in advancing linguistic analysis and natural language understanding. WSD plays a major role indirectly in various non-computing domains, not only in linguistically related domains. Garla and Brandt (2013) uses a knowledge-based WSD method for accurate clinical text classification. These studies have been further developed for clinical abbreviation disambiguation (Wu et al., 2015b,a) and acronyms disambiguation (Moon et al., 2015). The financial industry has mainly benefited from accurate WSD, primarily in stock price prediction (Hogenboom et al., 2021) and market prediction based on sentiment analysis of news headlines (Seifollahi and Shajari, 2019). In mathematics (Jiang et al., 2025) and social science, semantic disambiguation has been used for information discovery (Diamantini et al., 2015), while the impact of WSD on social media text analysis on micro posts is explored (Sumanth and Inkpen, 2015). WSD has also facilitated the disambiguation of complex terminology in law and medicine, ensuring clarity and precision in critical decision-making processes (Buitelaar et al., 2006). These examples highlight how WSD bridges linguistic research and practical applications, fostering innovation and efficiency across various fields.

7 Discussion and Concluding Remarks

Despite LLMs improving English WSD, low-resourced languages still have not enjoyed the same benefits. Building fine-grained WordNets for morphologically rich languages remains challenging due to required human involvement. Domain-specific variations pose difficulties because of divergent sense distributions. English WSD techniques show promising trends, with LLMs successfully leveraging contextual awareness and position-based gloss encoding systems enhancing performance. Future research should explore integrating knowledge graphs with LLMs to address less frequent word ambiguity. Multilingual WSD, especially for low-resource languages, requires further development of neural models beyond the currently dominant knowledge-based approaches. Key findings from our systematic literature review:

- Building high-performance WSD models remains a significant challenge.
- Verb disambiguation lags behind nouns and adjectives in accuracy.
- Less frequent senses remain difficult to disambiguate, requiring balanced training data.
- Low-resource language WSD needs further exploration through multilingual approaches.

Limitations

While these limitations do not significantly affect the overall findings of this review, we believe it is important to acknowledge them for transparency and clarity. The primary focus of this review was to discuss the challenges and limitations of current WSD methods in both English and non-English domains. However, greater emphasis has been placed on English WSD due to the wider availability of research in this area. The sample size for non-English WSD studies was limited by resource availability, and some relevant papers could not be included due to access restrictions. Additionally, extended abstracts and certain non-English studies were excluded due to the absence of sufficient results or details required for thorough analysis.

Ethics Statement

This paper has been conducted in compliance with the ethical standards of Anonymised University. The authors confirm that all sources have been appropriately cited and no conflicts of interest are related to this work. Generative AI tools were used solely to enhance the clarity and readability of the manuscript.

References

Reem Abdel-Salam. 2024. rematchka at arabicnlu2024: Evaluating large language models for arabic word sense and location sense disambiguation. In *Proceedings of The Second Arabic Natural Language Processing Conference*, pages 383–392.

Farah Adeeba and Sarmad Hussain. 2011. Experiences in building urdu wordnet. In *Proceedings of the 9th workshop on Asian language resources*, pages 31–35.

Eneko Agirre, Oier Lopez de Lacalle, Bernardo Magnini, Arantxa Otegi, German Rigau, and Piek Vossen. 2008. Semeval-2007 task 01: Evaluating wsd on cross-language information retrieval. In *Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007, Revised Selected Papers 8*, pages 908–917. Springer.

Eneko Agirre, Oier López De Lacalle, and Aitor Soroa. 2014. Random Walks for Knowledge-Based Word Sense Disambiguation. *Computational Linguistics*, 40(1):57–84.

Eneko Agirre and Aitor Soroa. 2009. Personalizing pagerank for word sense disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 33–41.

Elisabeth Aimelet, Veronika Lux, Corinne Jean, and Frédérique Segond. 1999. Wsd evaluation and the looking-glass. In *Proceedings of TALN*.

Boshra F Zopon Al-Bayaty and Shashank Joshi. 2016. Comparative analysis between naïve bayes algorithm and decision tree to solve wsd using empirical approach. *Lecture Notes on Software Engineering*, 4(1):82.

Marwah Alian and Arafat Awajan. 2020. Sense inventories for arabic texts. In *2020 21st International Arab Conference on Information Technology (ACIT)*, pages 1–4. IEEE.

Ahmed Aliwy, Ayad Abbas, and Ahmed Alkhayyat. 2021. Nerws: Towards improving information retrieval of digital library management system using named entity recognition and word sense. *Big Data and Cognitive Computing*, 5(4):59.

Janindu Arukgoda, Vidudaya Bandara, Samiththa Bashani, Vijayindu Gamage, and Daya Wimalasuriya. 2014. A word sense disambiguation technique for sinhala. In *2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology*, pages 207–211. IEEE.

Timothy Baldwin, Nam Kim Su, Francis Bond, Sanae Fujita, David Martinez, and Takaaki Tanaka. 2008. Mrd-based word sense disambiguation: Further extending lesk.

Mohamad Ballout, Anne Dedert, Nohayr Abdelmoneim, Ulf Krumnack, Gunther Heidemann, and Kai-Uwe Kühnberger. 2024. Fool me if you can! an adversarial dataset to investigate the robustness of lms in word sense disambiguation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 5042–5059.

Satanjeev Banerjee, Ted Pedersen, et al. 2003. Extended gloss overlaps as a measure of semantic relatedness. In *Ijcai*, volume 3, pages 805–810.

Edoardo Barba, Tommaso Pasini, and Roberto Navigli. 2021a. ESC: Redesigning WSD with Extractive Sense Comprehension. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4661–4672, Online. Association for Computational Linguistics.

Edoardo Barba, Luigi Procopio, and Roberto Navigli. 2021b. ConSeC: Word Sense Disambiguation as Continuous Sense Comprehension. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1492–1503, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Edoardo Barba, Luigi Procopio, and Roberto Navigli. 2022. ExtEnD: Extractive Entity Disambiguation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2478–2488, Dublin, Ireland. Association for Computational Linguistics.

Pierpaolo Basile, Annalina Caputo, and Giovanni Semeraro. 2014. An enhanced lesk word sense disambiguation algorithm through a distributional semantic model. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1591–1600.

Pierpaolo Basile, Lucia Siciliani, Elio Musacchio, and Giovanni Semeraro. 2025. Exploring the word sense disambiguation capabilities of large language models. *arXiv preprint arXiv:2503.08662*.

Osman Başkaya and David Jurgens. 2016. Semi-supervised learning with induced word senses for state of the art word sense disambiguation. *Journal of Artificial Intelligence Research*, 55:1025–1058.

Gábor Berend. 2020. Sparsity Makes Sense: Word Sense Disambiguation Using Sparse Contextualized Word Representations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8498–8508, Online. Association for Computational Linguistics.

Michele Bevilacqua, Roberto Navigli, et al. 2020. Breaking through the 80% glass ceiling: Raising the state of the art in word sense disambiguation by incorporating knowledge graph information. In *Proceedings of the conference-Association for Computational Linguistics. Meeting*, pages 2854–2864. Association for Computational Linguistics.

Michele Bevilacqua, Tommaso Pasini, Alessandro Raganato, and Roberto Navigli. 2021. Recent trends in word sense disambiguation: A survey. In *International joint conference on artificial intelligence*, pages 4330–4338. International Joint Conference on Artificial Intelligence, Inc.

Pushpak Bhattacharyya. 2017. *IndoWordNet*, pages 1–18. Springer Singapore, Singapore.

Terra Blevins, Mandar Joshi, and Luke Zettlemoyer. 2021. FEWS: Large-Scale, Low-Shot Word Sense Disambiguation with the Dictionary. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 455–465, Online. Association for Computational Linguistics.

Terra Blevins and Luke Zettlemoyer. 2020. Moving Down the Long Tail of Word Sense Disambiguation with Gloss Informed Bi-encoders. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1006–1017, Online. Association for Computational Linguistics.

Claudio Delli Bovi, Jose Camacho-Collados, Alessandro Raganato, and Roberto Navigli. 2017. Eurosense: Automatic harvesting of multilingual sense annotations from parallel text. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 594–600.

Melissa Bowerman. 1978. The acquisition of word meaning: An investigation into some current conflicts. In *The development of communication*, pages 263–287. Wiley.

Paul Buitelaar, Bernardo Magnini, Carlo Strapparava, and Piek Vossen. 2006. Domain-specific wsd. *Word Sense Disambiguation*, pages 275–298.

Andrei M Butnaru and Radu Tudor Ionescu. 2019. Shotgunwsd 2.0: An improved algorithm for global word sense disambiguation. *IEEE Access*, 7:120961–120975.

Andrei M Butnaru, Radu Tudor Ionescu, and Florentina Hristea. 2017. Shotgunwsd: An unsupervised algorithm for global word sense disambiguation inspired by dna sequencing. *arXiv preprint arXiv:1707.08084*.

Samuel Cahyawijaya, Ruochen Zhang, Holy Lovenia, Jan Christian Blaise Cruz, Hiroki Nomoto, and Alham Fikri Aji. 2024. Thank you, stingray: Multilingual large language models can not (yet) disambiguate cross-lingual word sense. *arXiv preprint arXiv:2410.21573*.

Jose Camacho-Collados, Claudio Delli Bovi, Alessandro Raganato, and Roberto Navigli. 2019. S ense d efs: a multilingual corpus of semantically annotated textual definitions: Exploiting multiple languages and resources jointly for high-quality word sense disambiguation and entity linking. *Language Resources and Evaluation*, 53:251–278.

Marine Carpuat and Dekai Wu. 2005. Evaluating the word sense disambiguation performance of statistical machine translation. In *Companion Volume to the Proceedings of Conference including Posters/Demos and tutorial abstracts*.

Marine Carpuat and Dekai Wu. 2007. Improving statistical machine translation using word sense disambiguation. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, pages 61–72.

Yee Seng Chan, Hwee Tou Ng, and David Chiang. 2007. Word sense disambiguation improves statistical machine translation. In *Proceedings of the 45th annual meeting of the association of computational linguistics*, pages 33–40.

Devendra Singh Chaplot and Ruslan Salakhutdinov. 2018. Knowledge-based Word Sense Disambiguation using Topic Models. ArXiv:1801.01900 [cs].

Ping Chen, Wei Ding, Chris Bowes, and David Brown. 2009. A fully unsupervised word sense disambiguation method using dependency knowledge. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 28–36.

Kenneth Ward Church. 2017. Word2vec. *Natural Language Engineering*, 23(1):155–162.

Paul Clough and Mark Stevenson. 2004. Cross-language information retrieval using eurowordnet and word sense disambiguation. In *European Conference on Information Retrieval*, pages 327–337. Springer.

Marta R Costa-Jussá and Mireia Farrús. 2014. Statistical machine translation enhancements through linguistic levels: A survey. *ACM Computing Surveys (CSUR)*, 46(3):1–28.

Debapratim Das Dawn, Abhinandan Khan, Soharab Hossain Shaikh, and Rajat Kumar Pal. 2023. A dataset for evaluating bengali word sense disambiguation techniques. *Journal of Ambient Intelligence and Humanized Computing*, 14(4):4057–4086.

Niladri Sekhar Dash, Pushpak Bhattacharyya, and Jyoti D Pawar. 2017. *The WordNet in Indian Languages*. Springer.

Robert David, Anna Kernerman, Ilan Kernerman, Nicolas Ferranti, and Assaf Siani. 2024. Multilingual word sense disambiguation for semantic annotations: Fusing knowledge graphs, lexical resources, and large language models.

Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22.

Claudia Diamantini, Alex Mircoli, Domenico Potena, and Emanuele Storti. 2015. Semantic disambiguation in a social information discovery system. In *2015 International Conference on Collaboration Technologies and Systems (CTS)*, pages 326–333. IEEE.

Asma Djaidri, Hassina Aliane, and Hamid Azzoune. 2023. The contribution of selected linguistic markers for unsupervised arabic verb sense disambiguation. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(8):1–23.

José Marcio Duarte, Samuel Sousa, Evangelos Milios, and Lilian Berton. 2021. Deep analysis of word sense disambiguation via semi-supervised learning and neural word representations. *Information Sciences*, 570:278–297.

Philip Edmonds and Scott Cotton. 2001. Senseval-2: overview. In *Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems*, pages 1–5.

Mohammed El-Razzaz, Mohamed Waleed Fakhr, and Fahima A Maghraby. 2021. Arabic gloss wsd using bert. *Applied Sciences*, 11(6):2567.

W Nelson Francis and Henry Kucera. 1979. Brown corpus manual. *Letters to the Editor*, 5(2):7.

Aparitosh Gahankari, Avinash S Kapse, Mohammad Atique, VM Thakare, and Arvind S Kapse. 2023a. Hybrid approach for word sense disambiguation in marathi language. In *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)*, pages 1–4. IEEE.

Aparitosh Gahankari, Dr Avinash S Kapse, Pranav K Biyani, and Dr Arvind S Kapse. 2023b. Marathi Word Sense Disambiguation using Bootstrapping Method. 10(10).

Chandra Ganesh, Sanjay K Dwivedi, Satya Bhushan Verma, and Manish Dixit. 2024. A systematic analysis of various word sense disambiguation approaches. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 13:e31602–e31602.

Vijay N Garla and Cynthia Brandt. 2013. Knowledge-based biomedical word sense disambiguation: an evaluation and application to clinical document classification. *Journal of the American Medical Informatics Association*, 20(5):882–886.

Alfio Gliozzo, Carlo Strapparava, and Ido Dagan. 2004. Unsupervised and supervised exploitation of semantic domains in lexical disambiguation. *Computer Speech & Language*, 18(3):275–299.

Arjun Gogoi and Nomi Baruah. 2022. A lemmatizer for low-resource languages: Wsd and its role in the assamese language. *Transactions on Asian and Low-Resource Language Information Processing*, 21(4):1–22.

Sharon Goldwater, Dan Jurafsky, and Christopher D Manning. 2010. Which words are hard to recognize? prosodic, lexical, and disfluency factors that increase speech recognition error rates. *Speech Communication*, 52(3):181–200.

Gurinder Pal Singh Gosal. 2015. A naive bayes approach for word sense disambiguation. *International Journal*, 5(7).

Daniel Guzman-Olivares, Lara Quijano-Sanchez, and Federico Liberatore. 2025. Sandwich: Semantical analysis of neighbours for disambiguating words in context ad hoc. *arXiv preprint arXiv:2503.05958*.

Dong Han, Junhui Li, Yachao Li, Min Zhang, and Guodong Zhou. 2019. Explicitly modeling word translations in neural machine translation. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 19(1):1–17.

Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. 2021. Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4):1–37.

Alexander Hogenboom, Alex Brojba-Micu, and Flavius Frasincar. 2021. The impact of word sense disambiguation on stock price prediction. *Expert systems with applications*, 184:115568.

Bairu Hou, Fanchao Qi, Yuan Zang, Xurui Zhang, Zhiyuan Liu, and Maosong Sun. 2020. Try to substitute: An unsupervised chinese word sense disambiguation method based on hownet. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1752–1757.

Florentina Hristea and Mihaela Colhon. 2020. The long road from performing word sense disambiguation to successfully using it in information retrieval: An overview of the unsupervised approach. *Computational Intelligence*, 36(3):1026–1062.

Luyao Huang, Chi Sun, Xipeng Qiu, and Xuanjing Huang. 2020. GlossBERT: BERT for Word Sense Disambiguation with Gloss Knowledge. ArXiv:1908.07245 [cs].

Nancy Ide, Collin F Baker, Christiane Fellbaum, and Rebecca J Passonneau. 2010. The manually annotated sub-corpus: A community resource for and by the people. In *Proceedings of the ACL 2010 conference short papers*, pages 68–73.

Vivek Iyer, Edoardo Barba, Alexandra Birch, Jeff Z Pan, and Roberto Navigli. 2023. Code-switching with word senses for pretraining in neural machine translation. *arXiv preprint arXiv:2310.14050*.

Areej Jaber and Paloma Martínez. 2021. Disambiguating clinical abbreviations using pre-trained word embeddings. In *Healthinf*, pages 501–508.

Goonjan Jain and DK Lobiyal. 2020. Word sense disambiguation using implicit information. *Natural Language Engineering*, 26(4):413–432.

Prajna Jha, Shreya Agarwal, Ali Abbas, and Tanveer Siddiqui. 2023a. Comparative analysis of path-based similarity measures for word sense disambiguation. In *2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP)*, pages 1–5. IEEE.

Prajna Jha, Shreya Agarwal, Ali Abbas, and Tanveer J Siddiqui. 2023b. A novel unsupervısed graph-based algorıthm for hindi word sense disambiguation. *SN Computer Science*, 4(5):675.

Prajna Jha, Shreya Agarwal, Ali Abbas, and Tanveer J. Siddiqui. 2023c. A Novel Unsupervısed Graph-Based Algorıthm for Hindi Word Sense Disambiguation. *SN Computer Science*, 4(5):675.

Shufan Jiang, Mary Ann Tan, and Harald Sack. 2025. Towards disambiguation of mathematical terms based on semantic representations. In *SCOLIA 2025-First International Workshop on Scholarly Information Access*.

Sanaa Kaddoura and Reem Nassar. 2024. Enhancedbert: A feature-rich ensemble model for arabic word sense disambiguation with statistical analysis and optimized data collection. *Journal of King Saud University-Computer and Information Sciences*, 36(1):101911.

Haoqiang Kang, Terra Blevins, and Luke Zettlemoyer. 2023a. Translate to Disambiguate: Zero-shot Multilingual Word Sense Disambiguation with Pretrained Language Models. ArXiv:2304.13803 [cs].

Haoqiang Kang, Terra Blevins, and Luke Zettlemoyer. 2023b. Translate to disambiguate: Zero-shot multilingual word sense disambiguation with pretrained language models. *arXiv preprint arXiv:2304.13803*.

In-Su Kang, Seung-Hoon Na, and Jong-Hyeok Lee. 2004. Influence of wsd on cross-language information retrieval. In *International Conference on Natural Language Processing*, pages 358–366. Springer.

Raihan Kibria, Sheikh Dipta, and Muhammad Adnan. 2024. On functional competence of llms for linguistic disambiguation. In *Proceedings of the 28th Conference on Computational Natural Language Learning*, pages 143–160.

Chandrakant D Kokane, Sachin D Babar, and Parikshit N Mahalle. 2021. An adaptive algorithm for lexical ambiguity in word sense disambiguation. In *Proceeding of First Doctoral Symposium on Natural Computing Research: DSNCR 2020*, pages 103–111. Springer.

SG Kolte and SG Bhirud. 2009. Exploiting links in wordnet hierarchy for word sense disambiguation of nouns. In *Proceedings of the International Conference on Advances in Computing, Communication and Control*, pages 20–25.

Neeraja Koppula, K. Srinivasa Rao, and B. VeeraSekharReddy. 2021. Word Sense Disambiguation Using Context Dependent Methods. In *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 1582–1590, Tirunelveli, India. IEEE.

Sawan Kumar, Sharmistha Jat, Karan Saxena, and Partha Talukdar. 2019. Zero-shot word sense disambiguation using sense definition embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5670–5681.

Andrey Kutuzov and Elizaveta Kuzmenko. 2019. To lemmatize or not to lemmatize: how word normalisation affects elmo performance in word sense disambiguation. *arXiv preprint arXiv:1909.03135*.

Rim Laatar, Chafik Aloulou, and Lamia Hadrich Belguith. 2023. Evaluation of Stacked Embeddings for Arabic Word Sense Disambiguation. *Computación y Sistemas*, 27(2).

Yurii Laba, Volodymyr Mudryi, Dmytro Chaplynskyi, Mariana Romanyshyn, and Oles Dobosevych. 2023. Contextual embeddings for ukrainian: A large language model approach to word sense disambiguation. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, pages 11–19.

Cuong Anh Le and Akira Shimazu. 2004. High wsd accuracy using naive bayesian classifier with rich features. In *Proceedings of the 18th Pacific Asia conference on language, information and computation*, pages 105–114.

Michael Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on Systems documentation*, pages 24–26.

Shou-de Lin and Karin Verspoor. 2008. A semantics-enhanced language model for unsupervised word sense disambiguation. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 287–298. Springer.

Ken Litkowski. 2004. Senseval-3 task: Automatic labeling of semantic roles. In *Proceedings of SENSEVAL-3, the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pages 9–12.

Daniel Loureiro and Alípio Jorge. 2019. Language modelling makes sense: Propagating representations through WordNet for full-coverage word sense disambiguation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5682–5691, Florence, Italy. Association for Computational Linguistics.

Daniel Loureiro, Kiamehr Rezaee, Mohammad Taher Pilehvar, and Jose Camacho-Collados. 2020. Language models and word sense disambiguation: An overview and analysis. *arXiv preprint arXiv:2008.11608*.

Fuli Luo, Tianyu Liu, Qiaolin Xia, Baobao Chang, and Zhifang Sui. 2018a. Incorporating Glosses into Neural Word Sense Disambiguation. ArXiv:1805.08028 [cs].

Fuli Luo, Tianyu Liu, Qiaolin Xia, Baobao Chang, and Zhifang Sui. 2018b. Incorporating glosses into neural word sense disambiguation. *arXiv preprint arXiv:1805.08028*.

Meng Lyu and Shasha Mo. 2023. Hsrg-wsd: A novel unsupervised chinese word sense disambiguation method based on heterogeneous sememe-relation graph. In *International Conference on Intelligent Computing*, pages 623–633. Springer.

Fabrizio Macagno and Sarah Bigi. 2018. Types of dialogue and pragmatic ambiguity. *Argumentation and Language—Linguistic, Cognitive and Discursive Explorations*, pages 191–218.

Maryellen C MacDonald, Neal J Pearlmutter, and Mark S Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological review*, 101(4):676.

Enrique Manjavacas and Lauren Fonteyn. 2022. Non-parametric word sense disambiguation for historical languages. In *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*, pages 123–134.

Vivek A Manwar, Rita L Gupta, and AB Manwar. 2024. Word sense disambiguation for marathi language in cross language information retrieval. *Recent Advancements in Science and Technology*, page 155.

Rui Mao, Kai He, Xulang Zhang, Guanyi Chen, Jinjie Ni, Zonglin Yang, and Erik Cambria. 2024. A survey on semantic processing techniques. *Information Fusion*, 101:101988.

Federico Martelli, Stefano Perrella, Niccolò Campolungo, Tina Munda, Svetla Koeva, Carole Tiberius, and Roberto Navigli. 2024. Dibimt: A gold evaluation benchmark for studying lexical ambiguity in machine translation. *Computational Linguistics*, pages 1–79.

Jorge Martinez-Gil. 2023a. Context-Aware Semantic Similarity Measurement for Unsupervised Word Sense Disambiguation. ArXiv:2305.03520 [cs].

Jorge Martinez-Gil. 2023b. Context-aware semantic similarity measurement for unsupervised word sense disambiguation. *arXiv preprint arXiv:2305.03520*.

Mosima Anna Masethe, Hlaudi Daniel Masethe, and Sunday O Ojo. 2024. Context-aware embedding techniques for addressing meaning conflation deficiency in morphologically rich languages word embedding: A systematic review and meta analysis. *Computers*, 13(10):271.

George A. Miller. 1992. WordNet: A lexical database for English. In *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, February 23-26, 1992*.

George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

George A Miller, Claudia Leacock, Randee Tengi, and Ross T Bunker. 1993. A semantic concordance. In *Human Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey, March 21-24, 1993*.

Anupam Mondal, Iti Chaturvedi, Dipankar Das, Rajiv Bajpai, and Sivaji Bandyopadhyay. 2015. Lexical resource for medical events: A polarity based approach. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 1302–1309.

Sungrim Moon, Bridget McInnes, and Genevieve B Melton. 2015. Challenges and practical approaches with word sense disambiguation of acronyms and abbreviations in the clinical domain. *Healthcare informatics research*, 21(1):35–42.

Andrea Moro and Roberto Navigli. 2015. Semeval-2015 task 13: Multilingual all-words sense disambiguation and entity linking. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 288–297.

Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics*, 2:231–244.

G Craig Murray and Rebecca Green. 2004. Lexical knowledge and human disagreement on a wsd task. *Computer Speech & Language*, 18(3):209–222.

Preethi Nanjundan and Eappen Zachariah Mathews. 2023. An Analysis of Word Sense Disambiguation (WSD). In Sarika Jain, Sven Groppe, and Nandana Mihindukulasooriya, editors, *Proceedings of the International Health Informatics Conference*, volume 990, pages 251–259. Springer Nature Singapore, Singapore. Series Title: Lecture Notes in Electrical Engineering.

Roberto Navigli, Michele Bevilacqua, Simone Conia, Dario Montagnini, Francesco Cecconi, et al. 2021. Ten years of babelnet: A survey. In *IJCAI*, pages 4559–4567. International Joint Conferences on Artificial Intelligence Organization.

Roberto Navigli, David Jurgens, and Daniele Vannella. 2013. Semeval-2013 task 12: Multilingual word sense disambiguation. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 222–231.

Roberto Navigli and Simone Paolo Ponzetto. 2010. Babelnet: Building a very large multilingual semantic network. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 216–225.

Steven Neale, Luís Gomes, Eneko Agirre, Oier Lopez de Lacalle, and António Branco. 2016. Word sense-aware machine translation: Including senses as contextual features for improved translation models. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2777–2783.

Cheng Niu, Wei Li, Rohini K Srihari, Huifeng Li, and Laurie Crist. 2004. Context clustering for word sense disambiguation based on modeling pairwise context similarities. In *Proceedings of SENSEVAL-3, the third international workshop on the evaluation of systems for the semantic analysis of text*, pages 187–190.

Riccardo Orlando, Simone Conia, Fabrizio Brignone, Francesco Cecconi, and Roberto Navigli. 2021. AMuSE-WSD: An All-in-one Multilingual System for Easy Word Sense Disambiguation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 298–307, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Miguel Ortega-Martín, Óscar García-Sierra, Alfonso Ardoiz, Jorge Álvarez, Juan Carlos Armenteros, and Adrián Alonso. 2023. Linguistic ambiguity analysis in ChatGPT. ArXiv:2302.06426 [cs].

Hirkani Padwad, Gunjan Keswani, Wani Bisen, Rajshree Sharma, Sopan Thakre, and Aditi Tiwari. 2024. Leveraging contextual factors for word sense disambiguation in hindi language. *International Journal of Intelligent Systems and Applications in Engineering*, 12(12s):129–136.

Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, et al. 2021. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, 372.

Alok Ranjan Pal, Diganta Saha, Niladri Sekhar Dash, and Antara Pal. 2018. Word sense disambiguation in bangla language using supervised methodology with necessary modifications. *Journal of The Institution of Engineers (India): Series B*, 99:519–526.

Martha Palmer, Hwee Tou Ng, and Hoa Trang Dang. 2006. Evaluation of wsd systems. *Word Sense Disambiguation: Algorithms and Applications*, pages 75–106.

Tommaso Pasini, Roberto Navigli, et al. 2017. Train-o-matic: Large-scale supervised word sense disambiguation in multiple languages without manual training data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 78–88.

Pasini, Tommaso and Navigli, Roberto. 2020. Train-O-Matic: Supervised Word Sense Disambiguation with no (manual) effort. *Artificial Intelligence*, 279:103–215.

Ted Pedersen. 2006. Unsupervised corpus-based methods for wsd. *Word sense disambiguation*, pages 133–166.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Sandun Sameera Perera, Lahiru Prabhath Jayakodi, Deshan Koshala Sumanathilaka, and Isuri Anuradha. 2025. Indonlp 2025 shared task: Romanized sinhala to sinhala reverse transliteration using bert. In *Proceedings of the First Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages*, pages 135–140.

Laura Plaza, Antonio J Jimeno-Yepes, Alberto Díaz, and Alan R Aronson. 2011. Studying the correlation between different word sense disambiguation methods and summarization effectiveness in biomedical texts. *BMC bioinformatics*, 12:1–13.

Sameer Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. 2007a. Semeval-2007 task-17: English lexical sample, srl and all words. In *Proceedings of the fourth international workshop on semantic evaluations (SemEval-2007)*, pages 87–92.

Sameer S Pradhan, Eduard Hovy, Mitch Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2007b. Ontonotes: A unified relational semantic representation. In *International Conference on Semantic Computing (ICSC 2007)*, pages 517–526. IEEE.

Xiao Pu, Nikolaos Pappas, James Henderson, and Andrei Popescu-Belis. 2018. Integrating weakly supervised word sense disambiguation into neural machine translation. *Transactions of the Association for Computational Linguistics*, 6:635–649.

Muhammad Qorib, Geonsik Moon, and Hwee Tou Ng. 2024. Are decoder-only language models better than encoder-only language models in understanding word meaning? In *Findings of the Association for Computational Linguistics ACL 2024*, pages 16339–16347.

Alessandro Raganato, Claudio Delli Bovi, and Roberto Navigli. 2016. Automatic construction and evaluation of a large semantically enriched wikipedia. In *IJCAI*, pages 2894–2900.

Alessandro Raganato, Jose Camacho-Collados, and Roberto Navigli. 2017. Word Sense Disambiguation: A Unified Evaluation Framework and Empirical Comparison. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 99–110, Valencia, Spain. Association for Computational Linguistics.

KM Tahsin Hassan Rahit, Khandaker Tabin Hasan, Md Al-Amin, and Zahiduddin Ahmed. 2018. Banglanet: Towards a wordnet for bengali language. In *Proceedings of the 9th Global WordNet Conference*, pages 1–9.

S Rajendran and KP Soman. 2017. Malayalam wordnet. *The WordNet in Indian Languages*, pages 119–145.

Sankaravelayuthan Rajendran, Selvaraj Arulmozi, B Kumara Shanmugam, S Baskaran, and S Thiagarajan. 2002. Tamil wordnet. In *Proceedings of the first international global WordNet conference. Mysore*, volume 152, pages 271–274.

Zhengfei Ren, Annalina Caputo, and Gareth Jones. 2024. A few-shot learning approach for lexical semantic change detection using gpt-4. In *Proceedings of the 5th Workshop on Computational Approaches to Historical Language Change*, pages 187–192.

Hossein Rouhizadeh, Mehrnoush Shamsfard, and Masoud Rouhizadeh. 2020. Knowledge Based Word Sense Disambiguation with Distributional Semantic Expansion for the Persian Language. In *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 329–335, Mashhad, Iran. IEEE.

Alexander James Rudnick. 2018. CROSS-LINGUAL WORD SENSE DISAMBIGUATION FOR LOW-RESOURCE HYBRID MACHINE TRANSLATION.

Ali Saeed, Rao Muhammad Adeel Nawab, Mark Stevenson, and Paul Rayson. 2019. A word sense disambiguation corpus for urdu. *Language Resources and Evaluation*, 53:397–418.

Kalyanamalini Sahoo and V Eshwarchandra Vidyasagar. 2003. Kannada wordnet-a lexical database. In *TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region*, volume 4, pages 1352–1356. IEEE.

Oscar Sainz, Oier Lopez de Lacalle, Eneko Agirre, and German Rigau. 2023. What do Language Models know about word senses? Zero-Shot WSD with Language Models and Domain Inventories.

KP Sruthi Sankar, PC Reghu Raj, and V Jayan. 2016. Unsupervised approach to word sense disambiguation in malayalam. *Procedia Technology*, 24:1507–1513.

Jumi Sarmah and Shikhar Kr Sarma. 2016. Word sense disambiguation for assamese. In *2016 IEEE 6th international conference on advanced computing (IACC)*, pages 146–151. IEEE.

Bianca Scarlini, Tommaso Pasini, and Roberto Navigli. 2019. Just "onesec" for producing multilingual sense-annotated data. In *Proceedings of the 57th annual meeting of the Association for Computational Linguistics*, pages 699–709.

Bianca Scarlini, Tommaso Pasini, and Roberto Navigli. 2020a. Sense-Annotated Corpora for Word Sense Disambiguation in Multiple Languages and Domains.

Bianca Scarlini, Tommaso Pasini, and Roberto Navigli. 2020b. Sense-annotated corpora for word sense disambiguation in multiple languages and domains. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5905–5911, Marseille, France. European Language Resources Association.

Bianca Scarlini, Tommaso Pasini, and Roberto Navigli. 2020c. Sensembert: Context-enhanced sense embeddings for multilingual word sense disambiguation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8758–8765.

Bianca Scarlini, Tommaso Pasini, Roberto Navigli, et al. 2020d. With more contexts comes better performance: Contextualized sense embeddings for all-round word sense disambiguation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3528–3539. The Association for Computational Linguistics.

Saeed Seifollahi and Mehdi Shajari. 2019. Word sense disambiguation application in sentiment analysis of news headlines: an applied approach to forex market prediction. *Journal of Intelligent Information Systems*, 52:57–83.

Chingakham Ponykumar Singh and H Mamata Devi. 2024. Additional diverse techniques for improvising lesk algorithm to enhance manipuri word sense disambiguation. *Nanotechnology Perceptions*, pages 171–190.

Satyendr Singh, Renish Rauniyar, and Murali Manohar. 2021. Nepali word-sense disambiguation using variants of simplified lesk measure. *Data Science: Theory, Algorithms, and Applications*, pages 41–57.

Benjamin Snyder and Martha Palmer. 2004. The english all-words task. In *Proceedings of SENSEVAL-3, the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pages 41–43.

Yang Song, Xin Cai Ong, Hwee Tou Ng, and Qian Lin. 2021. Improved Word Sense Disambiguation with Enhanced Sense Representations. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4311–4320, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Rhea Sukthanker, Soujanya Poria, Erik Cambria, and Ramkumar Thirunavukarasu. 2020. Anaphora and coreference resolution: A review. *Information Fusion*, 59:139–162.

Deshan Sumanathilaka, Nicholas Micallef, and Julian Hough. 2024a. Assessing gpt's potential for word sense disambiguation: A quantitative evaluation on prompt engineering techniques. In *2024 IEEE 15th Control and System Graduate Research Colloquium (ICSGRC)*, pages 204–209. IEEE.

Deshan Sumanathilaka, Nicholas Micallef, and Julian Hough. 2025. GlossGPT: GPT for Word Sense Disambiguation using Few-shot Chain-of-Thought Prompting. *Procedia Computer Science*.

Deshan Sumanathilaka, Nicholas Micallef, and Ruvan Weerasinghe. 2024b. Swa-Bhasha Dataset: Romanized Sinhala to Sinhala Adhoc Transliteration Corpus. In *2024 4th International Conference on Advanced Research in Computing (ICARC)*, pages 189–194, Belihuloya, Sri Lanka. IEEE.

T.G.D.K. Sumanathilaka, Nicholas Micallef, and Julian Hough. 2024c. Can LLMs assist with Ambiguity? A Quantitative Evaluation of various Large Language Models on Word Sense Disambiguation.

TGDK Sumanathilaka, Nicholas Micallef, and Julian Hough. 2024d. Can llms assist with ambiguity? a quantitative evaluation of various large language models on word sense disambiguation.

T.G.D.K. Sumanathilaka, Ruvan Weerasinghe, and Y.H.P.P. Priyadarshana. 2023. Swa-Bhasha: Romanized Sinhala to Sinhala Reverse Transliteration using a Hybrid Approach. In *2023 3rd International Conference on Advanced Research in Computing (ICARC)*, pages 136–141, Belihuloya, Sri Lanka. IEEE.

Chiraag Sumanth and Diana Inkpen. 2015. How much does word sense disambiguation help in sentiment analysis of micropost data? In *Proceedings of the 6th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 115–121.

Kaveh Taghipour and Hwee Tou Ng. 2015. One million sense-tagged instances for word sense disambiguation and induction. In *Proceedings of the nineteenth conference on computational natural language learning*, pages 338–344.

Carla Treloar, Sharon Champness, Paul L Simpson, and Nick Higginbotham. 2000. Critical appraisal checklist for qualitative research studies. *The Indian Journal of Pediatrics*, 67:347–351.

Nemika Tyagi, Sudeshna Chakraborty, Aditya Kumar, Nzanzu Katasohire Romeo, et al. 2022. Word sense disambiguation models emerging trends: a comparative analysis. In *Journal of Physics: Conference Series*, volume 2161, page 012035. IOP Publishing.

L Alfonso Urena-López, Manuel Buenaga, and Jose M Gomez. 2001. Integrating linguistic resources in tc through wsd. *Computers and the Humanities*, 35:215–230.

A Vaswani. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*.

Loïc Vial, Benjamin Lecouteux, and Didier Schwab. 2019. Sense vocabulary compression through the semantic knowledge of wordnet for neural word sense disambiguation. *arXiv preprint arXiv:1905.05677*.

RV Vidhu Bhala and S Abirami. 2014. Trends in word sense disambiguation. *Artificial Intelligence Review*, 42:159–171.

Jun Wang, Yanhui Li, Xiang Huang, Lin Chen, Xiaofang Zhang, and Yuming Zhou. 2023. Back deduction based testing for word sense disambiguation ability of machine translation systems. In *Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis*, pages 601–613.

Ming Wang and Yinglin Wang. 2020. A Synset Relation-enhanced Framework with a Try-again Mechanism for Word Sense Disambiguation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6229–6240, Online. Association for Computational Linguistics.

Ming Wang and Yinglin Wang. 2021. Word sense disambiguation: Towards interactive context exploitation from both word and sense perspectives. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5218–5229.

Yinglin Wang, Ming Wang, and Hamido Fujita. 2020. Word Sense Disambiguation: A comprehensive knowledge exploitation framework. *Knowledge-Based Systems*, 190:105030.

Viraj Welgama, Dulip Lakmal Herath, Chamila Liyanage, Namal Udalamatta, Ruvan Weerasinghe, and Tissa Jayawardana. 2011. Towards a sinhala wordnet. In *Proceedings of the Conference on Human Language Technology for Development*, volume 7.

Dominic Widdows and Beate Dorow. 2002. A graph model for unsupervised lexical acquisition. In *COLING 2002: The 19th International Conference on Computational Linguistics*.

Gregor Wiedemann, Steffen Remus, Avi Chawla, and Chris Biemann. 2019. Does bert make any sense? interpretable word sense disambiguation with contextualized embeddings. *arXiv preprint arXiv:1909.10430*.

Y Wu, JC Denny, ST Rosenbloom, RA Miller, DA Giuse, M Song, and Hua Xu. 2015a. A preliminary study of clinical abbreviation disambiguation in real time. *Applied clinical informatics*, 6(02):364–374.

Yonghui Wu, Jun Xu, Yaoyun Zhang, and Hua Xu. 2015b. Clinical abbreviation disambiguation using neural word embeddings. In *Proceedings of BioNLP 15*, pages 171–176.

Deyi Xiong and Min Zhang. 2014. A sense-based translation model for statistical machine translation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1459–1469.

Jung H Yae, Nolan C Skelly, Neil C Ranly, and Phillip M LaCasse. 2024. Leveraging large language models for word sense disambiguation. *Neural Computing and Applications*, pages 1–18.

Jung H Yae, Nolan C Skelly, Neil C Ranly, and Phillip M LaCasse. 2025. Leveraging large language models for word sense disambiguation. *Neural Computing and Applications*, 37(6):4093–4110.

Wenlin Yao, Xiaoman Pan, Lifeng Jin, Jianshu Chen, Dian Yu, and Dong Yu. 2021. Connect-the-Dots: Bridging Semantics between Words and Definitions via Aligning Word Sense Inventories. ArXiv:2110.14091 [cs].

Hee Suk Yoon, Eunseop Yoon, John Harvill, Sunjae Yoon, Mark Hasegawa-Johnson, and Chang D. Yoo. 2022. SMSMix: Sense-Maintained Sentence Mixup for Word Sense Disambiguation. ArXiv:2212.07072 [cs].

Yue Yu, Simiao Zuo, Haoming Jiang, Wendi Ren, Tuo Zhao, and Chao Zhang. 2021. Fine-Tuning Pre-trained Language Model with Weak Supervision: A Contrastive-Regularized Self-Training Approach. ArXiv:2010.07835 [cs].

Deping Zhang, Zhaohui Yang, Xiang Huang, and Yanhui Li. 2025. Extensive mutation for testing of word sense disambiguation models. *Information and Software Technology*, page 107734.

Xuefeng Zhang, Richong Zhang, Xiaoyang Li, Fanshuang Kong, Junfan Chen, Samuel Mensah, and Yongyi Mao. 2023. Word Sense Disambiguation by Refining Target Word Embedding. In *Proceedings of the ACM Web Conference 2023*, pages 1405–1414, Austin TX USA. ACM.

Zhi Zhong and Hwee Tou Ng. 2012. Word sense disambiguation improves information retrieval. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 273–282.

Qiaoli Zhou, Gu Yue, and Yuguang Meng. 2019. Incorporating hownet-based semantic relatedness into chinese word sense disambiguation. In *Workshop on Chinese Lexical Semantics*, pages 359–370. Springer.

Shi Zong, Ashutosh Baheti, Wei Xu, and Alan Ritter. 2022. Extracting a knowledge base of COVID-19 events from social media. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3810–3823, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

A Appendix

The keywords used for the study

For general ambiguity : Ambiguity Resolution in NLP, Linguistic Ambiguity, Syntactic Ambiguity, Semantic Ambiguity, Pragmatic Ambiguity, Polysemy and Homonymy, Ambiguity Detection, Contextual Disambiguation, Discourse Ambiguity, Structural Ambiguity, Ambiguity in Sentence Parsing, Parsing Strategies for Ambiguity, Multi-interpretation in Language, Word Ambiguity in Text, Ambiguity in Machine Translation.

For WSD: Word Sense Disambiguation (WSD) Algorithms, Supervised Word Sense Disambiguation, Unsupervised Word Sense Disambiguation, Semi-supervised Word Sense Disambiguation, Weakly Supervised WSD, Neural Networks for

WSD, WSD in Natural Language Processing (NLP), Contextual Embeddings for WSD, BERT for Word Sense Disambiguation, Knowledge-based WSD, Dictionary-based WSD, Lexical Semantics in WSD, Machine Learning for WSD, Sense Inventory for WSD, Sense Representation in WSD, Graph-based Approaches in WSD, Evaluation Metrics for WSD, Ambiguity Resolution in NLP, Contextualized Word Representations, WSD Applications in Information Retrieval, Cross-lingual WSD, Hybrid Approaches to WSD, Explainable WSD Models, Multi-sense Embedding Models, Deep Learning for WSD, Comparative Studies in WSD Algorithms, Lesk Algorithm for WSD, Translation-based WSD, WSD in Machine Translation, Biomedical Word Sense Disambiguation, Chinese Word Sense Disambiguation, Arabic Word Sense Disambiguation, Hindi Word Sense Disambiguation, Co-occurrence Graphs for WSD, Mesh Indexing for WSD, Capsule Networks for WSD, Self-Attention Mechanisms in WSD, Language Models for WSD, Generative Adversarial Networks (GANs) for WSD, Accuracy in Word Sense Disambiguation, Short Literature Review on WSD, Context Exploitation for WSD.

Sub-heading used for extracting information

The labelling used to extract the information for the papers is shown in the table 2.

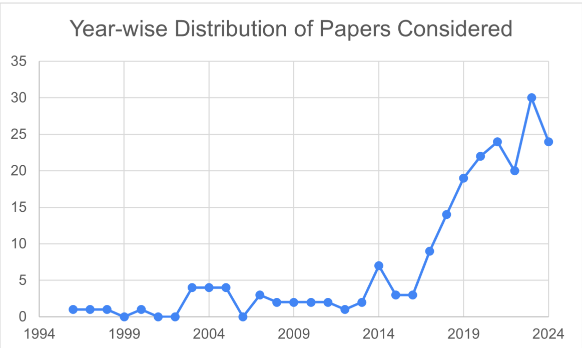

Figure 3: Year-wise paper distribution (1995-2024)

Selected Attribute	**Description**
Year	Indicate the year of publication
Overview	A summary of the paper
Technology stack	List of technologies used including models, programming languages and specific modules
Improvements to the existing algorithms	Advancement on current literature and changes done on algorithms
Contributions to the State of Art	major contributions of the study focusing on performance and resource creations
Methodology	The steps/ algorithm used to solve the problem
Limitations	Identified issues by the original author and the annotator
Dataset used	list of datasets used for training and testing
Evaluation Matrix	Quantitative and Qualitative evaluation techniques used
Usage/benefits	Additions indirect outcomes
Keywords	tags to easily filter the literature
Challenges	Identified challenges during the study period

Table 2: Summary of selected attributes and their descriptions

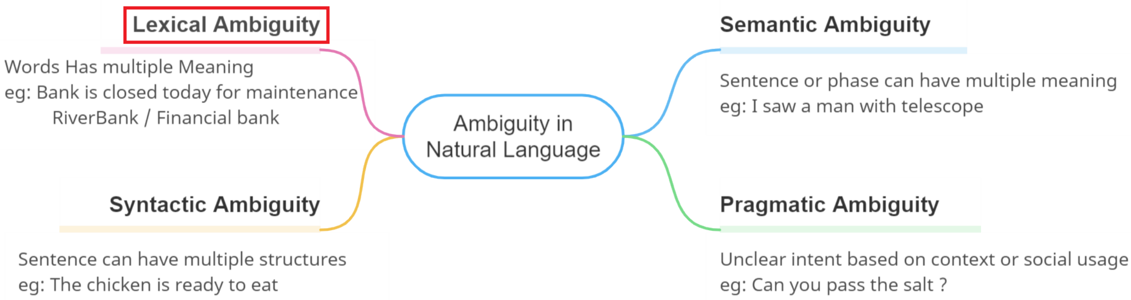

Figure 4: Ambiguities in Natural Text

Model	**Supervision**	**BERT/LMs**	**Gloss Use**	**Sense Embeddings**	**Contextualized Embeddings**	**Nearest Neighbor**	**Advanced LM***
GAS (2018)	Hybrid	✗	✓	✗	✗	✗	✗
EWISE (2019)	Hybrid	✓	✗	✗	✓	✗	✗
Vial et al. (2019)	Supervised	✓	✗	✓	✓	✗	✓
LMMS (2019)	Knowledge-based	✓	✗	✓	✓	✓	✓
GlossBERT (2020)	Supervised	✓	✓	✗	✓	✗	✓
ARES (2020)	Semi-supervised	✓	✗	✓	✓	✓	✓
SenseEMBERT (2020)	Hybrid	✓	✗	✓	✓	✓	✓
EWISER (2020)	Supervised	✓	✗	✗	✓	✗	✓
SEMEQ (2021)	Hybrid	✓	✓	✗	✓	✗	✓
ESR (2021)	Supervised	✓	✗	✗	✓	✗	✓
CONSEC (2021)	Supervised	✓	✗	✗	✓	✗	✓
SACE (2021)	Supervised	✓	✗	✗	✓	✗	✓
ESCHER (2021)	Supervised	✓	✓	✗	✓	✗	✓
RTWE (2023)	Supervised	✓	✗	✗	✓	✗	✓
GlossGPT (2024)	Hybrid	✓	✓	✗	✓	✗	✓

Table 3: Comparative A nalysis on papers of Meta analysis. LM: Language models *Architectural enhancements or fine-tuning employed

Category	Dataset Description	Citation(s)
English	SemCor – Manually annotated corpus from the Brown Corpus with 226K annotations.	(Miller et al., 1993; Francis and Kucera, 1979)
	MASC-WSA – 45 lemma-pos pairs with crowd-sourced annotations.	(Ide et al., 2010)
	Princeton WordNet Gloss Corpus – 330,499 manually/semi-automatically annotated glosses.	(Miller, 1995; Baldwin et al., 2008)
	OntoNotes – Rich syntactic and semantic annotations across genres.	(Pradhan et al., 2007b)
	FEWS – Corpus from Wiktionary with 121k sentences and 35k polysemous words.	(Blevins et al., 2021)
	OMSTI – Semi-automated corpus from English-Chinese parallel data for bilingual WSD.	(Taghipour and Ng, 2015)
	SEW (Semantically Enriched Wikipedia) – Propagated annotations across Wikipedia using links.	(Raganato et al., 2016)
	FOOL – Four different test sets for evaluating WSD model robustness.	(Ballout et al., 2024)
Multilingual	BabelNet – Integrates lexicographic and encyclopedic knowledge for 263 languages.	(Navigli et al., 2021; Navigli and Ponzetto, 2010)
	SenseDefs – Automatic disambiguation of definitions in 263 languages based on Princeton Gloss Corpus.	(Camacho-Collados et al., 2019)
	EuroSense – Parallel corpus-based multilingual dataset for 21 languages using Europarl.	(Bovi et al., 2017)
	Train-o-Matic – Automatically annotated dataset without relying on parallel corpora.	(Pasini et al., 2017; Pasini, Tommaso and Navigli, Roberto, 2020)
	OneSeC – Domain-specific multilingual corpus.	(Scarlini et al., 2019, 2020b)
	DiBIMT – Benchmark dataset for 8 languages.	(Martelli et al., 2024)
Low-Resource Languages*	Bengali WSD Dataset – 100 polysemous words with 10 sense paragraphs each.	(Das Dawn et al., 2023; Pal et al., 2018)
	IndoWordNet – Lexical database for 18 Indian languages highlighting synsets and semantic relations.	(Dash et al., 2017; Bhattacharyya, 2017)
	Manually created Assamese corpus	(Sarmah and Sarma, 2016)
	Sinhala, Tamil, Malayalam, Urdu, BanglaNet, Assamese and Kannada WordNets Regional lexical databases for low-resource languages	(Welgama et al., 2011; Rajendran et al., 2002; Rajendran and Soman, 2017; Adeeba and Hussain, 2011; Saeed et al., 2019; Rahit et al., 2018; Sarmah and Sarma, 2016; Sahoo and Vidyasagar, 2003)
	Romanised Sinhala Dataset – For transliteration disambiguation.	(Sumanathilaka et al., 2024b)
	HowNet-based Chinese WSD dataset	(Zhou et al., 2019)

Table 4: Key WSD Datasets by Language Scope *Many studies related to low resource uses custom corpora